**Visão Geral Técnica: Modelos de Processamento de Áudio para Extração de Metadados**

**Introdução**

Este documento apresenta uma visão geral técnica de três modelos desenvolvidos para processar arquivos de áudio e extrair metadados relevantes. O objetivo principal destes modelos é automatizar a extração de informações cruciais de gravações de áudio, como dados de mercado, datas de pesquisa, nomes de clientes e pontuações NPS (Net Promoter Score).

**Objetivos Comuns dos Modelos**

1. Transcrever o conteúdo de arquivos de áudio para texto.
2. Extrair metadados específicos do texto transcrito.
3. Gerar um dataset estruturado com as informações extraídas.
4. Processar múltiplos arquivos de áudio em lote.

**Descrição dos Modelos**

**Modelo 1: LeituraAudios\_ModeloTreino01**

Este modelo utiliza a biblioteca SpeechRecognition com o reconhecedor do Google para transcrição de áudio e a biblioteca spaCy para processamento de linguagem natural (NLP).

**Características principais:**

* Transcrição de áudio usando Google Speech Recognition.
* Análise de entidades nomeadas com spaCy para extração de metadados.
* Processamento de arquivos MP3 e WAV.

**Modelo 2: LeituraAudios\_ModeloTreino02**

O segundo modelo emprega a biblioteca Whisper da OpenAI para transcrição de áudio e NLTK (Natural Language Toolkit) para processamento de linguagem natural.

**Características principais:**

* Transcrição de áudio offline usando Whisper.
* Extração de palavras-chave e análise de entidades com NLTK.
* Potencial para maior precisão em transcrições complexas.

**Modelo 3: LeituraAudios\_ModeloTreino03**

Este modelo adota uma abordagem mais simples, utilizando SpeechRecognition para transcrição e técnicas básicas de processamento de texto para extração de metadados.

**Características principais:**

* Transcrição de áudio usando Google Speech Recognition.
* Extração de metadados baseada em padrões de texto e frequência de palavras.
* Implementação mais leve e com menos dependências.

**Principais Diferenças entre os Modelos**

1. **Tecnologia de Transcrição:**
   * Modelo 1 e 3: Google Speech Recognition (online)
   * Modelo 2: Whisper (offline)
2. **Processamento de Linguagem Natural:**
   * Modelo 1: spaCy (avançado)
   * Modelo 2: NLTK (intermediário)
   * Modelo 3: Técnicas básicas de processamento de texto
3. **Complexidade e Dependências:**
   * Modelo 1: Complexidade média, requer spaCy
   * Modelo 2: Maior complexidade, requer Whisper e NLTK
   * Modelo 3: Menor complexidade, dependências mínimas
4. **Precisão Esperada:**
   * Modelo 1: Alta precisão em extração de entidades
   * Modelo 2: Potencialmente maior precisão em transcrições
   * Modelo 3: Precisão moderada, dependente da qualidade do áudio

**Código de Teste Comparativo**

Para avaliar e comparar o desempenho destes três modelos, foi desenvolvido um script de teste que executa cada modelo em um conjunto de arquivos de áudio e compara seus resultados.

**Funcionalidades do Código de Teste:**

1. Carrega dinamicamente cada modelo.
2. Executa cada modelo no mesmo conjunto de dados de áudio.
3. Mede o tempo de execução de cada modelo.
4. Coleta estatísticas básicas dos datasets gerados (número de linhas, colunas).
5. Salva os datasets gerados por cada modelo.
6. Compila e salva os resultados comparativos.

**Métricas de Comparação:**

* Tempo de execução
* Número de linhas processadas
* Colunas geradas no dataset
* Consistência dos dados extraídos

A análise comparativa destes modelos fornece insights valiosos sobre suas performances relativas, permitindo uma seleção informada do modelo mais adequado para diferentes cenários de aplicação no processamento de áudio e extração de metadados.